Mobile AR Using Pre-captured Omnidirectional Images

Takayuki Akaguma* Fumio Okura† Tomokazu Sato‡ Naokazu Yokoya§
Nara Institute of Science and Technology (NAIST)

Abstract

In the field of augmented reality (AR), geometric and photometric registration is routinely achieved in real time. However, real-time geometric registration often leads to misalignment (e.g., jitter and drift) due to the error from camera pose estimation. Due to limited resources on mobile devices, it is also difficult to implement state-of-the-art techniques for photometric registration on mobile AR systems. In order to solve these problems, we developed a mobile AR system in a significantly different way from conventional systems. In this system, captured omnidirectional images and virtual objects are registered geometrically and photometrically in an offline rendering process. The appropriate part of the pre-rendered omnidirectional AR image is shown to a user through a mobile device with online registration between the real world and the pre-captured image. In order to investigate the validity of our new framework for mobile AR, we conducted experiments using the prototype system on a real site in Todai-ji Temple, a famous world cultural heritage site in Japan.

CR Categories: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality;

Keywords: augmented reality, mobile system, omnidirectional image

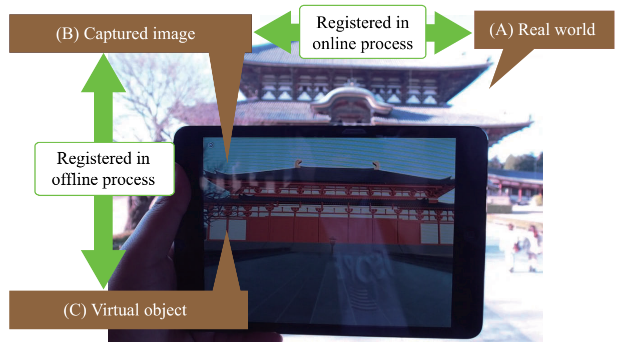

Figure 1: *Mobile AR using pre-generated images. (C) is superimposed on (B) in the offline process. Misalignment is expected to occur between (A) and (B). However, we argue that it has little influence on user experience.*

1 Introduction

Augmented reality (AR) using mobile devices (e.g., smartphones and tablet) can be used for applications such as the visualization of lost buildings at cultural heritage sites [Zoellner et al. 2009]. In order to realize highly immersive AR, it is necessary to achieve geometric registration, defined as the geometric alignment between real world and virtual objects, and photometric registration, that is, the consistency in appearance including lighting and shadowing. In most AR applications, geometric and photometric registration should be done in real time. Basically, geometric registration is achieved by estimating the pose (position and orientation) of the cameras used for virtual object superimposition. [Azuma 1997] [Tenmoku et al. 2004] [Hollerer et al. 1999] employed methods for sensor-based camera pose estimation using the GPS, compass, gyroscope, and accelerometer on the mobile device, and [Klein and Murray 2009] [Ventura and Hollerer 2012] [Taketomi et al. 2011] proposed vision-based methods using captured images. However, as shown in Figure 1, these real-time methods often lead to misalignment (e.g., jitter and drift) between the captured image (B) and the virtual object (C) due to the error in camera pose estimation that could detract from the immersion and naturalness of AR. On the other hand, photometric registration needs to estimate the light source from the captured image (B) and render the virtual object (C) using the light source. Due to limited resources on mobile devices, it is difficult to apply state-of-the-art techniques [Gruber et al. 2012] [Kán and Kaufmann 2012] [Lensing and Broll 2012] to photometric registration on mobile AR systems.

We propose a significantly different approach compared to conventional registration techniques. In the proposed approach, the captured image (B) and the virtual object (C) are registered geometrically and photometrically with high quality, even though misalignment occurs between the real world (A) and the captured image (B). That is, real scenes are captured beforehand, and the pre-captured image (B) and the virtual object (C) are registered geometrically and photometrically in an offline rendering process. The appropriate part of the pre-rendered AR image is presented to a user through a mobile device based on pose information acquired from the mobile device in real time.

Unlike conventional system using images captured in real time for (B), the proposed system achieves

- geometric registration without any perceptible misalignment between the captured images (B) and the virtual objects (C) and
- high-quality photometric registration by offline rendering for virtual objects

with low computational cost on the mobile device. Although geometric and photometric misalignments between the real world (A) and the captured image (B) are larger than conventional systems, we expect that the impact of the misalignments between (A) and (B) are much smaller than that of (B) and (C) in actual user experience. This is because users are used to accommodating for such mis-

*e-mail: takayuki-a@naist.jp
†e-mail: fumio-o@naist.jp
‡e-mail: tomoka-s@naist.jp
§e-mail: yokoya@naist.jp

alignment through ordinary mobile AR or traditional digital photographic interfaces that arise from the misalignment between the center of the device's camera and the human eye. To investigate the effects of this misalignment on users, we conducted experiments at a real site in Todai-ji Temple (hereafter, Todaiji), a famous world cultural heritage site in Japan, using a prototype system based on the proposed framework.

2 Mobile AR Using Pre-captured Omnidirectional Images

Unlike conventional mobile AR, where all registration processes are done in real time, our proposed method is divided into online and offline processes. In the offline process, images are captured by an omnidirectional camera at several fixed points and the virtual object is rendered with high quality in the offline process using real-world lighting. In the online process, the system shows the user perspective images generated from the omnidirectional AR image based on the estimated pose of the mobile device.

2.1 Offline Process: Capturing Omnidirectional Images and Generating Pre-rendered Images

In the offline process, virtual objects are superimposed onto omnidirectional images captured at fixed locations. First, the position and orientation of the fixed camera should be estimated, e.g., by using visual landmarks in the captured image. In order to achieve high-quality photometric registration between the virtual objects and pre-captured images, the virtual objects are rendered by image-based lighting (IBL) [Debevec 1998] using light maps estimated from the pre-captured image. It should be noted that the following two problems arise if we simply use a captured raw image (see Figure 2) for the light map and the background image:

1. The luminance is inaccurate for IBL light maps due to the limited dynamic range of the camera.

2. The locations of moving objects (e.g., person's figure) in the pre-captured image are different from the real world in real time. It could make users feel uncomfortable.

To solve the first problem, we employed the method in [Okura et al. 2012] for generating omnidirectional high dynamic range (HDR) images from multiple omnidirectional images captured with multiple exposure values of the camera. To solve the second problem, moving objects are removed from omnidirectional images. Moving objects are simply removed by using the time-series median of pixel values on the omnidirectional image sequences captured at fixed points [Arai et al. 2010]. The remaining objects after median filtering are removed by image inpainting [Kawai et al. 2009]. An example of an omnidirectional HDR image after moving object removal is shown in Figure 3.

The virtual objects are then superimposed on the pre-captured images after the moving objects have been removed. Here, IBL, a time-consuming but high-quality rendering method, uses the pre-captured omnidirectional HDR images to add photorealistic lighting and shadowing effects to the virtual objects. Note that the proposed framework is not restricted to specific methods for the offline rendering and that the rendered images can be manually edited. Finally, tone mapping [Reinhard et al. 2002], a luminance compression process for HDR images, generates the final omnidirectional images, as shown in Figure 4.

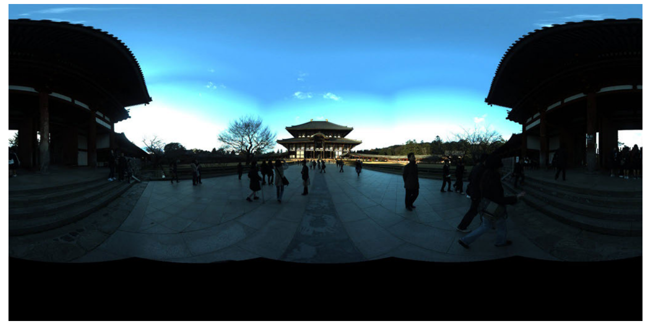

Figure 2: *Image captured by omnidirectional camera.*

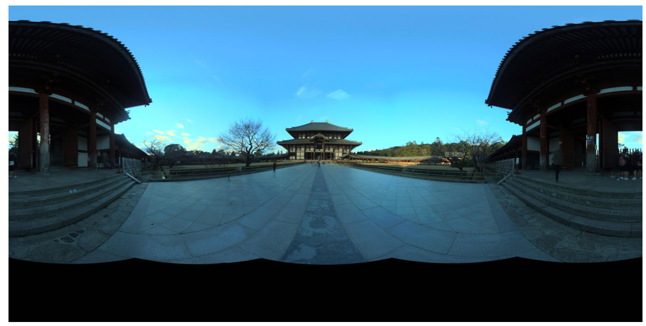

Figure 3: *Omnidirectional HDR image where person's figure have been removed. (Tone mapped by [Reinhard et al. 2002])*

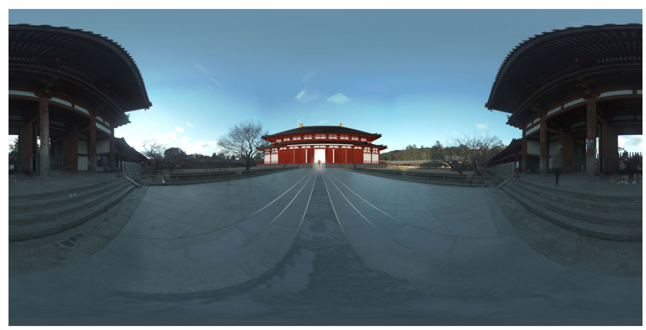

Figure 4: *Omnidirectional image augmented by IBL and manual post-editing.*

2.2 Online Process: Presenting AR Images to a User through a Mobile Device

In the online process, the position and orientation of the mobile device are acquired by attached sensors: GPS, compass, gyroscope, and accelerometer. The pre-captured point closest to the position of the device is selected. Then the appropriate part of the pre-rendered omnidirectional AR image is converted to a perspective image based on the device orientation and presented to the user through the mobile display.

3 Application for Virtual History Experience

We developed an application based on the proposed framework for a virtual history experience at the site in Todaiji, and conducted experiments with the application and the public. In these experiments, computer graphics (CG) contents consisting of ancient architecture

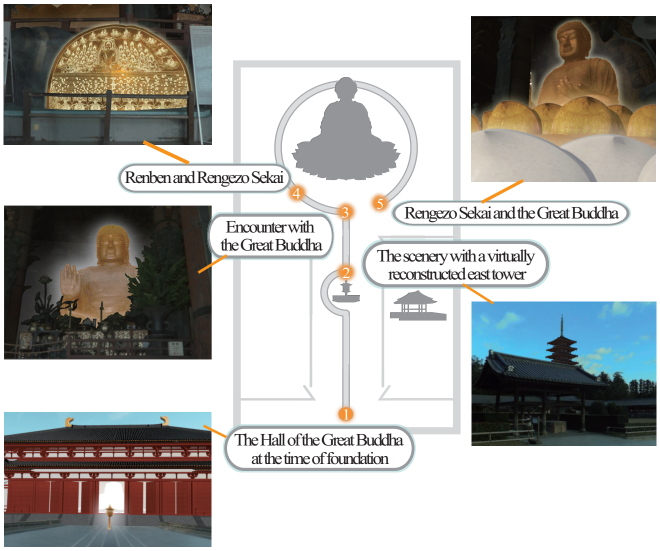

Figure 5: *Content presented by the application in the Hall of the Great Buddha.*

and heirloom quality statues lost to history [1] were superimposed in the offline process on the omnidirectional images.

3.1 Application Interface

The application for the virtual history experience has two functions. One is a navigation function that shows the map of the Hall of the Great Buddha (see center of Figure 5) and the position of a user. When the user reaches one of the pre-captured points, the system switches to AR mode, and the user can see the AR contents through the display.

3.2 Contents Presented by the Application in the Hall of the Great Buddha

In the application for the virtual history experience, different content was prepared for five locations (see Figure 5). Details of each scene are described below.

1. The Hall of the Great Buddha at the time of foundation: The Hall of the Great Buddha in the Tempyo Period, said to be larger than the present one, is superimposed on the position of the present hall.
2. The scenery with a virtually reconstructed east tower: A miniature model of Todaiji at the time of foundation is exhibited in the Hall of the Great Buddha. The CG model of the east tower [2] is superimposed on the scene of its ruins.
3. Encounter with the Great Buddha: Although there are various opinions on the ancient appearance of the Great Buddha, in this application, the Great Buddha in the scene is covered with gold, as illustrated in an ancient picture scroll.
4. Renben and Rengezo Sekai: The Renben is a decoration of the pedestal of the Great Buddha. In the current era, only a replica exists. In the application, the replica glows gold with its detailed textures.

[1]This content was created by a team at the Graduate School of Media Design, Keio University.

[2]This content was created with reference to the miniature model with architectural validation by Yumiko Fukuda, Hiroshima Institute of Technology, and patina expression technology validation by Takeaki Nakajima, Hiroshima City University.

5. Rengezo Sekai and the Great Buddha: The Great Buddha and its pedestal are superimposed with decorations based on historical sources.

4 Public Experiments Using the Prototype System

4.1 Overview of Experiments

We investigated the usefulness of the application based on the proposed framework for virtual history experiences. The application was implemented on our prototype system that consisted of an iPad2, iPad (4th generation), and iPad mini (Apple Inc.), all running iOS6. Ladybug3 (Point Grey Research Inc.), an omnidirectional multi-camera unit, was used to capture the omnidirectional images. While the omnidirectional images had been captured on a fine day, the public experiments were carried out under a cloudy sky. In order to enhance the user experiences of the application, some sound effects were added. The pre-rendered images were generated by commercially available rendering software, 3ds Max and Maya (Autodesk, Inc.), as well as by manual post-editing. Participants browsed the AR images within two or three meters of the captured point of the omnidirectional image. Forty-six members of the public, whose ages are summarized in Table 1, answered questions after experiencing the application.

We asked the participants the following four questions by questionnaire.

Q1) Was this application useful for understanding the ancient Great Buddha and other historical objects?

Q2) Did you worry about the geometric misalignment between the CG and the scenes displayed in the device?

Q3) Did you worry about the geometric misalignment between the scenes in the display and the scene you saw with your eyes?

Q4) Did you worry about the difference of color between the scenes in the display and the scene you saw with your eyes?

Q1 evaluated the usefulness of the application. The participants chose one of five degrees of usefulness. Q2 evaluated the quality of registration between (B) and (C) in Figure 1. Q3 and Q4 are questions about registration the between (A) and (B). The participants chose one of five degrees of naturalness for Q2 to Q4.

4.2 Result and Discussion

Figure 6 shows the results for Q1. About 90 % of the participants answered "very useful" or "useful" for Q1. This result indicates that the proposed system, a mobile AR system using pre-captured omnidirectional images, was useful with respect to experiencing history.

Figure 7 shows the results of the questionnaire for Q2 to Q4. Focusing on geometric registration, over 70 % of the participants answered "very natural" or "natural" for Q2, concerning the alignment between the virtual objects and the pre-captured images, as well as Q3, concerning the alignment between the pre-captured image and the real world. This result indicates that most of the participants did not perceive the geometric misalignment of the proposed system when there was small discrepancy (i.e., within two or three meters) between the camera position of the pre-captured image and the mobile device. In our experimental situation, the photometric

Table 1: *Age composition of participants*

Age	20-29	30-39	40-49	50-59	60-69	70-
Number of people	5	4	6	12	12	7

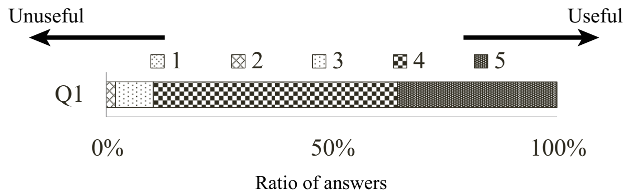

Figure 6: *Stacked bar graph of the degree of usefulness. (Q1)*

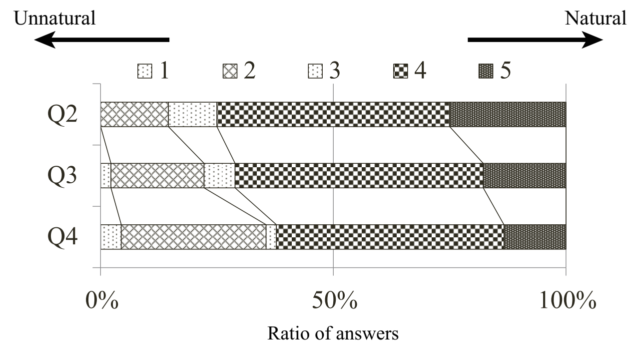

Figure 7: *Stacked bar graph of the degree of naturalness. (Q2-Q4)*

misalignment between the real world and the pre-captured image was more noticeable according to the results of Q4. About 40 % of the participants gave an "unnatural" or "very unnatural" assessment for Q4. Thus, photometric registration is essential for giving a more natural experience within our framework.

5 Conclusion and Future Work

This paper described a novel framework for mobile augmented reality. The framework is divided into two steps. First, the captured omnidirectional image and the virtual object are registered geometrically and photometrically in an offline rendering process. Next, the appropriate part of the pre-rendered omnidirectional AR image is presented to the user through a mobile device based on its orientation. The results of a questionnaire for given to participants in an experiment in Todaiji demonstrated that the proposed system was useful for experiencing history. Through the public experiment, we strongly believe that our framework could greatly benefit actual AR applications because of its stability and robustness. However, we also confirmed that the photometric misalignment between the real world and the captured images looked unnatural to users. In future work, we will improve the quality of the photometric registration between the real world and the captured images by updating the pre-captured images depending on various factors such as the weather and time.

Acknowledgements

This research was partially supported by Grant-in-Aids for Scientific Research (A), No. 23240024, the "Ambient Intelligence" project sponsored by MEXT, and NAIST Advanced Research Partnership Project. We thank members of NAIST-Keio joint research group. We also thank Todaiji for giving us the opportunity to conduct public experiments.

References

ARAI, I., HORI, M., KAWAI, N., ABE, Y., ICHIKAWA, M., SATONAKA, Y., NITTA, T., NITTA, T., FUJII, H., MUKAI, M., ET AL. 2010. Pano Umechika: A crowded underground city panoramic view system. In *Proc. 7th Int'l Symp. Distributed Computing and Artificial Intelligence (DCAI'10)*. 173–180.

AZUMA, R. T. 1997. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments 6*, 4, 355–385.

DEBEVEC, P. 1998. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proc. ACM SIGGRAPH'98*, 189–198.

GRUBER, L., RICHTER-TRUMMER, T., AND SCHMALSTIEG, D. 2012. Real-time photometric registration from arbitrary geometry. In *Proc. 11th IEEE Int'l Symp. Mixed and Augmented Reality (ISMAR'12)*, 91–97.

HOLLERER, T., FEINER, S., AND PAVLIK, J. 1999. Situated documentaries: Embedding multimedia presentations in the real world. In *Proc. 3rd IEEE Int'l Symp. Wearable Computers (ISWC'99)*, 79–86.

KÁN, P., AND KAUFMANN, H. 2012. High-quality reflections, refractions, and caustics in augmented reality and their contribution to visual coherence. In *Proc. 11th IEEE Int'l Symp. Mixed and Augmented Reality (ISMAR'12)*, 99–108.

KAWAI, N., SATO, T., AND YOKOYA, N. 2009. Image inpainting considering brightness change and spatial locality of textures and its evaluation. In *Proc. 3rd Pacific-Rim Symp. Image and Video Technology (PSIVT'09)*, 271–282.

KLEIN, G., AND MURRAY, D. 2009. Parallel tracking and mapping on a camera phone. In *Proc. 8th IEEE Int'l Symp. Mixed and Augmented Reality (ISMAR'09)*, 83–86.

LENSING, P., AND BROLL, W. 2012. Instant indirect illumination for dynamic mixed reality scenes. In *Proc. 11th IEEE Int'l Symp. Mixed and Augmented Reality (ISMAR'12)*, 109–118.

OKURA, F., KANBARA, M., AND YOKOYA, N. 2012. Full spherical high dynamic range imaging from the sky. In *Proc. 9th IEEE Int'l Conf. Multimedia and Expo (ICME'12)*, 325–332.

REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. In *Proc. ACM SIGGRAPH'02*, 267–276.

TAKETOMI, T., SATO, T., AND YOKOYA, N. 2011. Real-time and accurate extrinsic camera parameter estimation using feature landmark database for augmented reality. *Int'l Journal of Computers and Graphics 35*, 4, 768–777.

TENMOKU, R., KANBARA, M., AND YOKOYA, N. 2004. A positioning method combining specification of users absolute position and dead reckoning for wearable augmented reality systems. In *Proc. 3rd CREST/ISWC Workshop on Advanced Computing and Communicating Techniques for Wearable Information Playing*, 19–22.

VENTURA, J., AND HOLLERER, T. 2012. Wide-area scene mapping for mobile visual tracking. In *Proc. 11th IEEE Int'l Symp. Mixed and Augmented Reality (ISMAR'12)*, 3–12.

ZOELLNER, M., KEIL, J., DREVENSEK, T., AND WUEST, H. 2009. Cultural heritage layers: Integrating historic media in augmented reality. In *Proc. 15th Int'l Conf. Virtual Systems and Multimedia (VSMM'09)*, 193–196.